DOCUMENT RESUME

ED 247 291 TM 840 471

AUTHOR Harrison, Colin
TITLE Criteria for Evaluating Microcomputer Software for Reading Development: Observations Based on Three British Case Studies.
PUB DATE Apr 84
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; Check Lists; *Courseware; *Evaluation Criteria; Evaluation Methods; Foreign Countries; *Microcomputers; *Reading Instruction
IDENTIFIERS Context Input Process Product Evaluation Model; England

ABSTRACT

Differences in sets of criteria for evaluating microcomputer software are discussed. They are set against the results of three studies in which teachers in the United Kingdom evaluated five programs which were used in reading or English lessons. A comparison of the checklist criteria with the case study data was made using Stake's (1967) matrix of evaluation concerns. This suggested a heavy emphasis on antecedents in the checklists and on transactions in the case studies. In general, neither checklists nor case studies devoted great attention to empirically measured outcomes. A possible interpretation of the results is that while the checklists focussed on intrinsic evaluation, the case studies themselves focussed on practical classroom issues, notably attention and motivation. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247291

Criteria for evaluating microcomputer software for reading development: observations based on three British case studies.

COLIN HARRISON

Nottingham University, UK

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. Harrison

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper contributed to Critique Session Design and Evaluation in Instructional Development, AERA Annual Meeting, New Orleans 1984.

Colin Harrison
School of Education
University Park
NOTTINGHAM NG7 2RD
England

TM 810 471

Criteria for evaluating microcomputer software for reading development: observations based on three British case studies.

Abstract

Differences in sets of criteria for evaluating microcomputer software are discussed. They are set against the results of three studies in which UK teachers evaluated five programs which were used in reading or English lessons. A comparison of the checklist criteria with the case study data was made using Stake's (1967) matrix of evaluation concerns. This suggested a heavy emphasis on antecedents in the checklists and on transactions in the case studies. In general, neither checklists nor case studies devoted great attention to empirically measured outcomes. A possible interpretation of the results is that while the checklists focussed on intrinsic evaluation, the case studies themselves focussed on practical classroom issues, notably attention and motivation.

Criteria for evaluating microcomputer software for reading development: observations based on three British case studies.

The problem

According to Lathrop (1982), the critical evaluation of educational microcomputer programs in the US has not kept pace with the proliferation of software packages, with reviews of less than 5 percent appearing in print. A further problem surrounds the issue of what criteria should be adopted for evaluation. An examination of five recently published sets of guidelines for software evaluation (Jelden, 1981; Golub, 1982; Devall, 1983; Burkhardt et al., 1982; Adams and Jones, 1983) reveals that a number of different assumptions are made by specialists on both sides of the Atlantic about what questions a software review should address.

Aims of the study

This paper sets out to compare the evaluative assumptions built into published sets of guidelines with those derived more directly from three small UK case studies of microcomputer usage in reading/language classes. The case studies provide data on teacher and student reaction to five computer programs, each of which was used in a small-group context by children in the 9-13 age range.

The need for a conceptual framework for comparing and analysing the guidelines

As Robert Stake has pointed out (1980), oversimplification obfuscates. Nevertheless, in seeking to compare five very different sets of guidelines, some procedure for data reduction is essential. In this case it is proposed to use a variant of Stake's own description-judgment matrix (1967, 1980) in order to structure an analysis of the content of the lists. Stake originally offered his matrix as an aid to evaluators who were devising a "shopping list" of what data to gather, and its seems worthwhile to apply it retrospectively in order to analyse and compare the issues and concerns which are implicit in the checklists in the present study. This analysis will be of interest in illuminating some of the areas of emphasis and omission in the five sets of guidelines, but the Stake matrix will also be used for an analysis of the data of the three case-study reports. The data collection and reporting for the case studies was carried out for the most part by non-specialists in evaluation. A comparison of the two matrices will therefore provide an indication of the extent to which there is a match between the issues and concerns in the guidelines and those which surface in the classroom.

The original matrix consisted of a four-by-three array of cells: the horizontal axis was labelled intents, observations, standards and judgments, while the vertical axis was labelled antecedents, transactions and outcomes. Stake also had a thirteenth free-floating box labelled rationale. The horizontal axis was

divided, with intents and observations labelled as the description matrix, and standards and judgments labelled as the judgment matrix. This division suggested that Stake saw a sharp distinction between the concepts of description and observation. In fact, he acknowledged in his later paper (1980) that while he felt that the matrix could still be useful in planning an evaluation, the concept of observations is an extremely broad one, and could, in certain circumstances, encompass intents, standards, judgments and statements of rationale.

In the present paper, two sub-divisions of Stake's categories, intents and standards have been omitted. This has been done partly for clarity of presentation, but there is also evidence that these two categories are relatively minor in comparison with the categories of observations and judgments. In Clift's (1981) study of checklists for whole-school evaluation, for example, observations and judgments accounted for 142 out of the 156 items recorded using the full matrix.

It is perhaps worthwhile to offer a brief gloss on how the categories have been interpreted, since the issue of interpretation is both subjective and problematical. Antecedents, transactions and outcomes have been interpreted as relating to observations or judgments which are made respectively before, during or after classroom activity. This might seem the obvious interpretation, but in fact others are possible: for example, the question 'Do the children work in groups while using this program?' might seem to be a straightforward question of observing a transaction in the

classroom. However, to a curriculum developer, having the children work in groups might be a desired outcome of the use of the program. In the present analysis, however, the term outcomes is restricted to post hoc data, collected after the classroom session has ended. Equally, if an observation or judgment can be made before the classroom session begins, it will be classified as an antecedent. Thus a question such as 'Are there no more than three frames before a call for a response?' would be classed as an antecedent observation, since it could be answered in advance of the session by the teacher alone.

The question of differentiating between observations and judgments can also be problematical. In many cases there is little doubt: 'Are supplementary materials provided?' would seem to be a question which can be resolved unequivocally by examining the package- an observation. In contrast, a question such as 'Is the program free of pedagogical errors?' is hardly an issue which can be decided uneqivocally through description or observation; it would therefore be classed as a judgment.

As an example of a more difficult question to classify, one could consider the following: 'Is the program logically crashproof?' In this case, the teacher might try to answer the question by testing the program before the lesson. He or she might find a bug which causes the program to crash when certain keys are depressed- the question is unequivocally resolved- an observation, therefore. Suppose, however, no bug is found. In this case one could argue that the teacher has to make a judgment, and that the question is analoguous to 'Are all

possible user errors trapped and help messages provided?', which would certainly seem to be a difficult question to answer unequivocally.

Perhaps the best solution to this problem would be to accept that the notions of observation and judgment are not dichotomous, but rather regions at opposite ends of a continuum. Thus, while there is bound to be a subjective element in classifying questions as matters of observation or judgment, it is only in the middle of the continuum that that subjectivity will lead to unreliable judgments, and this need not therefore invalidate the whole decision-making procedure.

An analysis of the five checklists

The five checklists described below were found as a result of a survey of the educational computing literature made in England in 1983. The provenance of the checklists varied. The Adams and Jones list (1983, pp.129-131) is given at the end of a book on the place of the microcomputer in the humanities curriculum, and follows a statement in which the authors freely give their opinions on which educational publishers are producing worthwhile software and support materials. Burkhardt et al (1982, pp.85-94), by contrast, take a much less partisan view, and offer their checklist as part of an in-service pack designed to help teachers become more systematic evaluators of their own practice. The book emphasises the use of the microcomputer, but much of it would be appropriate for supporting formative and summative evaluation of other types of teaching material.

Of the three US checklists, two appeared in widely-circulated journals. Devall's list (1983, p.553) appeared as an open letter in the Journal of Reading, while that of Golub (1982, pp. 28-29) appeared as part of an article in The Computing Teacher. Finally, Jelden's list (1983, p. 159) was reprinted from another source as part of an extensive annotated bibliography in a specialist book for reading teachers on computer applications in their subject.

The items in the checklists were assigned to Stake's categories in the manner described above, and the result is shown in Table 1 (see Appendix A for an annotated example of one of the checklists). While it would be inappropriate to analyse the data too finely, a number of points may be made about differences between

[Table 1 about here]

the checklists. Firstly, there is an overall weight of emphasis which in terms of number of items gives

antecedents > transactions > outcomes.

Secondly, there is an overall emphasis, especially marked in the two UK studies, of

judgments > observations.

Looking more closely at the US lists, it is interesting to note the similarity between the lists of Devall and Golub. Jelden, by contrast, provides the only example of a checklist in which observations outnumber judgments.

What do these differences suggest in practical terms? In general, the emphasis on antecedent judgments perhaps reflects a wish to encourage an intrinsic evaluation of the educational

goals of the software, and to address pedagogical considerations such as whether the content is clearly organised and presented. The emphasis on antecedent observations, the second largest category overall, perhaps reflects a concern with technical considerations concerning the mechanics of use.

In some respects, this emphasis on antecedents is hardly surprising. Teachers usually have to make judgments about the likely worth of a program before they actually have an opportunity to try it out in the classroom. Generally speaking, it is not commercially viable to make inspection copies of software available: procedures for unlocking 'protected' software become common knowledge too rapidly. The authors of the checklists will have been well aware of this, and their guidelines therefore make few assumptions about the possibility of any classroom-based evaluation. This offers a pragmatic explanation for the emphasis on antecedents. We shall return later in this paper to the issue of precisely what interpretation should be put on an apparent lack of attention to trasnsactions and outcomes in four of the five checklists. Before that, however, it seems best to introduce and describe the main data source in this report, the three case studies. This will enable a contrastive account to be attempted, and will permit a fuller discussion of the applicability of Stake's matrix.

The three case studies

Case Study 1 was a dissertation completed as part of an in-service B.Ed. degree (Chan, 1983). It was based on an evaluation of two reading development programs, STORYBOARD and CLUES, both of which feature word deletion as a means of encouraging attentive reading and group discussion. STORYBOARD gives a totally deleted text, and information is available from prior exposure to the passage and from proportional length blanks which are given complete with punctuation; CLUES is a cloze-type exercise of the more familiar variety. In a crossover design, two groups of 6 students aged thirteen worked with both programs, using one of two specially selected short stories on each of the programs. Their responses and reactions were recorded on sound tape during and after the two sessions of activity, and the students also completed a questionnaire and cloze reading comprehension post-tests.

Case Study 2 reports the use of "Adventure Game" programs and an arcade game similar to "Pac-Man" in English lessons with a class of 25 twelve- to thirteen-year-olds. Over two six-week periods the students worked in small groups to produce either creative writing or a guide for other students who might wish to learn the strategies of each game. Two teachers worked with the class, and they kept a written record of their evaluation of the students' use of the microcomputers.

Case Study 3 reports the results of a formative evaluation of WILT, a spelling game which gives students information about likely letter patterns in English. The program contains a matrix of bigram frequencies derived from an analysis of the prose of newspapers and novels; the student can call up

histograms showing how likely it is that any letter of the alphabet will be followed by any other. Data collection was carried out in six schools, three near London and three close to Nottingham. Data collection was based on classroom observation, a questionnaire for teachers, a discussion with children and with the teachers, and unsolicited verbal or written comment.

These were rather different types of study, and in seeking to systemise an approach to applying Stake's categories one faces some problems. In the event, Case Studies 2 and 3 were not too difficult to analyse: they each amounted to less than twelve pages of text, and a statement-by-statement rating of comments was not onerous.

Case Study 1, by contrast, was much more problematical. At which points in a dissertation can one be said to locate the statements which most define the concerns of the study? This was especially difficult in the present case since the whole topic was on the theme of evaluation. One obvious candidate for analysis would presumably be the hypothesis section. On investigation, however, it was clear that there was a slight discontinuity between what was actually studied in some depth and what was highlighted in the hypotheses. Chan's hypotheses stressed those issues which were tested through cloze and reading comprehension, but they did not emphasise her interest in the transactions of the classroom, nor her intention to administer an attitude questionnaire. By contrast, however, in a section titled Introduction and statement of the problem, Chan does give a list of the questions which the study attempts to explore, and this includes reference to both the quantitative and qualitative

facets of her work. Another section of the study which gives an indication of her interests as an evaluator is the appendix, which includes a transcript of an interview with a group of children about the positive and negative aspects of using microcomputers in school.

After further consideration, therefore, it was decided to focus solely on these two aspects of Chan's study for the Stake analysis. In making this decision it was recognised that the issue of selection is complex, and one which might well have been approached differently. Thus, although her study totalled 70 pages plus appendices, in the present analysis it yielded only nineteen items which were categorised using the Stake matrix.

Appendix B gives an example of material from one of the case studies, together with an indication of how the statements were classified.

Results of analysis of case study data

The results of applying Stake's categories to the data in the case studies are shown in Table 2. As has already been noted, the decision to focus on two relatively limited sections of Chan's dissertation explains the comparatively small number of

(Table 2 about here)

items relating to Case Study 1. The data for Case Studies 2 and 3 are based on pooled results for two and six respondents respectively, and it is perhaps worth noting that although the

individual results are not shown, there were in fact fairly similar distributions within each of the two groups.

The main emphases shown in Table 2 are in the areas of transaction observations, transaction judgments, and outcome judgments. Together these account for 247 out of 295 statements analysed. Transaction observations were generally descriptions of student activity, such as 'pupils paid much more attention to the letter count' (Case Study 3, Respondent 4), or of teacher activity: 'I opened my mouth to shout "Tracey, don't shout!" but the word "Down!" came out instead.' (Case Study 2, Respondent 1). Transaction judjments were generally opinions which led to tactical decisions during the lessons, or which were aspects of a formative evaluation of a program in action: 'their enthusiasm was also noticeable and they needed a teacher to keep control' (Case Study 3, Respondent 1); 'They were just beginning to make interesting moves when their time was up' (Case Study 2, Respondent 2). Outcome judgments were generally part of a summative evaluation of the program, lesson, or associated coursework. These were opinions which were not substantiated by corroborative evidence: 'I thought that it (a piece of written work) lacked a certain realistic quality.' (Case Study 2, Respondent 2); 'With more appropriate words, I see no reason why less able readers and younger children should not be able to use the program beneficially.' (Case Study 3, Respondent 6).

After these three categories, the next largest is that of antecedent judgments. In Case Study 1, the issues which were assigned to this category were all culled from the interview section in the appendix, e.g.: 'Do you think if you have learned

to use the computer at school it will be useful to you when you leave?'; 'Do you think both boys and girls should learn to use the computer?'. In the other case studies, too, antecedent judgments tended to highlight issues related to intrinsic evaluation: 'There is a danger that explicit language programs will lead back to the formal arid drill and practice language exercises which have now fallen into disrepute.' (Case Study 3, Respondent2); 'I consider it to be a very nice program which seems to retain interest and has a true educational value.' (Case Study 3, Respondent 3).

It is perhaps interesting to note that despite the practical nature of the three studies, the emphasis on empirically-determined outcome data was patchy. The fact that there were only two items in Case Study 1 which came into the outcomes observation category should not be taken to imply that post-test results were unimportant: in fact these points were the central questions about the relationship between reading on the microcomputer and gains in comprehension. In Case Study 3, however, not a single reference is made by any of the six respondents to any empirical examination of whether children learn to spell by using the program WILT. It is as if attention was focussed solely on intuitive assessments of motivation and task-oriented activity, which together with a consideration of the program's implicit educational philosophy formed the basis of the final evaluation.

A comparison of the checklists and case studies

The aim of this paper is to compare the evaluative assumptions

built into the five sets of guidelines with those distilled from the three case studies, and it is now possible to offer some comment on the differences between the two, drawing initially upon apparent differences in emphasis which are suggested by the Stake matrix analysis. For convenience, the totals of Tables 1 and 2 have been reproduced alongside eachother in Table 3, and the results expressed in percentage form.

(Table 3 about here)

The most striking difference between the two sets of items in Table 3 is perhaps the relative salience of antecedents. If these are represented as they were earlier in terms of greatest to least, the following pattern emerges:

Checklists- antecedents > transactions > outcomes

Case Studies- transactions > outcomes > antecedents

Antecedents shift from the dominant to the least dominant category, while in both groups transactions attract more attention than outcomes. So far as the observation-judgment continuum is concerned, judgments tend to outnumber observations in both checklists and case studies, with the exceptions of Jelden's checklist and the transactions section of Case Study 2.

Discussion

What do the kind of differences shown up in Table 3 relate to in real terms? Do the differences in emphasis between the guidelines and the checklists imply importantly different evaluative perspectives, or are the differences mere artefacts,

created by the application of some rather arbitrary decision procedures on a singularly amorphous set of data? It has already been admitted that there is subjectivity in the application of the Stake matrix to any dataset, but it has equally been argued that this need not invalidate its use. It has also been pointed out that a strict quantitative approach to the numerical data would be inappropriate: the two occurences of outcome observation items in Case Study 3 referred to aspects of that study to which a great deal of attention was given. To apply inferential non-parametric statistics to this data would therefore be potentially misleading. Nevertheless, there remain a number of points which emerge from the comparison of the checklists and case studies, and which are well worth consideration despite these caveats. To emphasise their tentativeness, the points will be expressed as questions:

Why do antecedents dominate the checklists?

Is this an inevitable result of an agenda-setting operation?

If it is, then why do Burkhardt et al have so many items in other categories?

What is the significance of the apparent subordination of antecedents in the case studies?

Does this suggest an inattention to issues of intrinsic evaluation, or is attention to those issues masked by the crudeness of the matrix analysis?

What is the significance of the apparent inattention to empirically-determined outcomes in both checklists and case studies?

The fact that Table 3 raises all these questions may in itself be regarded as important. Quantification is not valuable in absolute terms, but only insofar as it performs a useful data reduction function, and draws attention to trends and patterns. In the present study, the type of material analysed included lists, segments of oral discourse, a teacher's lesson journal and a formal evaluation report submitted to a publisher. These data are very different, and not easily compared one with another without some systematic basis for analysis.

A possible interpretation of the emphasis on observational and judgmental antecedents in the evaluation guidelines might be that teachers are enjoined to consider the classroom potential of the software in terms of its mechanics of use and also its intrinsic educational merit. Equally, a possible interpretation of the emphases in the case studies on observational and judgmental transactions, and on judgmental outcomes might be that when teachers evaluate material, their attention is directed by the exigencies of the classroom towards immediate and pragmatic concerns. In such conditions, concerns such as time on task, student motivation and cooperation are likely to be much more dominant than either long-term pedagogical or philosophical issues.

These interpretations are open to debate, but it is suggested that they are important enough to merit further discussion, and if the Stake analysis has helped to point up the issue, it has perhaps served a useful purpose.

Conclusions

In the present study, sets of theoretical guidelines for focussing on evaluation issues have been compared with with the results of three practical studies in which evaluation issues are foregrounded and explored. It has been suggested that some potentially important differences of emphasis have emerged, and that in facilitating such comparisons, an analysis based on Stake's (1967) matrix can be of value, provided that it is used with circumspection.

In an area which is expanding so rapidly as that of microcomputers in education there is an urgent need not only for evaluation, but for the assumptions built into evaluations and evaluation guidelines to be made clear. The results of this study suggest that the criteria of theoreticians and practitioners may differ in important ways, and that these possible differences should be further explored.

References

Adams A. and Jones. E. (1983) Teaching Humanities in the Microelectronic Age Milton Keynes Open University Press, pp. 129-131.

Burkhardt, H. and Fraser, R., with Clowes, M., Eggleston, J. and Wells, C. (1982) Design and Development of Programs as Teaching Material London: Council for Educational Technology. pp. 85-94.

Chan, R. Y. P. (1983) Microcomputers and reading development. Unpublished B.Ed. Dissertation, University of Nottingham.

Devall, Y. (1983) Evaluating microcomputer software for reading instruction. Journal of Reading. 26. No. 6. p.553.

Golub, L.S. (1982) Some criteria for selecting microcomputer courseware for the teaching of English. The Computing Teacher. 10, No. 2. pp.28-29.

Jelden, D.L. (1981) The microcomputer as a multi-user interactive instructional system. AEDS Journal 14, No. 4. pp. 208-217.

Lathrop, A. (1982) Microcomputer software for instructional use: where are the critical reviews? The Computing Teacher, 9, No. 6, pp.22-26.

Stake, R.E. (1967) The countenance of educational evaluation. Teachers College Record, 68, 523-40.

Stake, R.E. (1980) Program evaluation, particularly responsive evaluation. In W.B.Dockrell and David Hamilton (Eds) Rethinking Educational Research. London: Hodder and Stoughton, 72-87.

Table 1. Stake's categories for evaluation applied to five checklists for evaluating microcomputer software.

	Antecedents		Transactions		Outcomes	
	Obs.	Judg.	Obs.	Judg.	Obs.	Judg.
Adams and Jones	xxxxx	xxxxxxxxxx xxxxxx	x	xxxxx		x
Burkhardt et al.	xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxx	xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx x	xxxxxxxxxx xxxxxxxxx	xxxxxxxxxx xxxxxxxxxx xxxxxxxx	x	xx
Devall	xxxxxxx	xxxxxxxxx				
Golub	xxxxxx	xxxxxxxx				
Jeluen	xxxxxxxxxx xxxxx	xxx				
Totals						
UK	53	116	20	43	1	3
USA	28	20	0	0	0	0
Overall	81	136	20	43	1	3

Table 2. Stake's categories for evaluation applied to three case studies involving the evaluation of microcomputer software in schools.

	Antecedents		Transactions		Outcomes	
	Obs.	Judg.	Obs.	Judg.	Obs.	Judg.
Case Study 1		xxxxxx		xx	xx	xxxxxxxxx
Case Study 2		xxxxxxx	xxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxxxxx	xxxxxxxxx xxx	xxxxxxxxxx	xxxxxxxxx xxxxxxxxx xxxxx
Case Study 3	xx	xxxxxxxxx xxxxxxxxx x	xxxxxxxxx xxxxxxxxx xxxxxxxx	xxxxxxxxx xxxxxxxxx xxxxxxxxx xxx		xxxxxxxxx xxxxxxxxx xxxxxxxxx xxxxxx
Totals	2	34	88	68	12	91

Table 3. Comparison of totals of evaluation observations and judgments in checklists and case studies.

	Antecedents		Transactions		Outcomes		
	Obs.	J.	Obs.	J.	Obs.	J.	Total
Checklists	81	136	20	43	1	3	284
%	29	48	7	15	0	1	100
Case Studies	2	34	88	68	12	91	295
%	1	11	30	23	4	31	100

Appendix A

Sample checklist with Stake categories added

I. Content

i	Is the content of educational value?	AJ
ii	Is the content up to date and accurate?	AJ
iii	Is there provision for adding and updating content material?	AO

II. Technical Considerations

i	Is the package compatible with the computer(s) in use and any peripherals that are also needed?	AO
ii	Is the program easy to load and does it run immediately on loading?	AO
iii	Is the program capable of being used by students independently of a teacher?	AJ
iv	Is the program reliable and 'crash proof' in normal use?	AJ

III. Pedagogical Considerations

i	Is the purpose of the program clearly defined? Is it clear to the students?	AJ, TJ
ii	Does the program allow students to enter it at a variety of starting points at different levels?	AJ
iii	Is the presentation of the content:	
	clear	AJ
	logical	AJ
	consistent?	AJ
iv	Is there appropriate use of:	
	colour	AJ
	sound	AJ
	graphics?	AJ
	Are the computer and VDU being used to handle colour, sound and graphics appropriately in the classroom situation?	TJ
v	Does the program provide diagnostic help so as to suggest further appropriate activities for the student?	AJ

IV. Student Appeal and 'User-Friendliness'

i	Is the program motivating to the age range(s) for which it is intended?	OJ
ii	Does the program allow for student interaction and/or creativity?	OJ
iii	Is the program one that gives the student adequate and early feedback about progress?	TJ
iv	Does information about student error lead to 'prompts' so that the student can continue to proceed successfully with the program?	TO
v	Can the student easily exit from the program so as to avoid the frustration resulting from continued failure?	TJ

BEST COPY AVAILABLE

Appendix B

Sample from Case Study 3 with Stake Categories added

1.	2-4 per group. If 4 then one person tends to take charge. 2 tends to be a better number.	TO TJ
2.	Disks more reliable than cassettes.	AJ
3.	Cost is critical.	AJ
4.	Must be flexible.	AJ
5.	M.Ellis had to explain to his class basic hangman clues. Every word has a vowel etc. This could possibly go to the teacher's notes.	TO OJ
6.	The vocabulary is slightly too difficult for this 10/11 year old group. However, they did seem to be coping quite well. Must have the facility to put in your own vocabulary and to link it with your own reading schemes should this be desired.	TJ TJ OJ
7.	The computer is going to be in the classroom, and use in it so noise can get grating. Can it be turned down?	OJ
8.	Word score confusing. This needs to relate to the words tried by any one person. Also, it is a cumulative scheme, and doesn't reflect the word just done. For example the scoring can go 100%, 0%, 50%, 66%...	TJ
9.	Letter score is useful, and reflects the pupils facility with words. It would be useful to have some feedback. However, get away from percents and be far simpler. Say 'number of words tried?' and ' number of words achieved'. Also, for the letter count, this is better expressed as number of letters in the word 'is' and number of letters tried 'is'.	OJ
10.	Histographs are not always helpful/relevant. It would be more valuable to get children to pick out patterns in the English language. For example, what letters are likely to go with 'ia', 'ai', 'ci', 'ti', 'th',prefixes and suffixes.	OJ
11.	Children use a dictionary to help with words. The teacher here found follow-up word to find out the meaning useful.	TO TJ